

Two-sided Grassmann-Rayleigh quotient iteration*

P.-A. Absil[†]

P. Van Dooren[†]

Submitted for publication on 19 Apr 2007

Abstract

The two-sided Rayleigh quotient iteration proposed by Ostrowski computes a pair of corresponding left-right eigenvectors of a matrix C . We propose a Grassmannian version of this iteration, i.e., its iterates are pairs of p -dimensional subspaces instead of one-dimensional subspaces in the classical case. The new iteration generically converges locally cubically to the pairs of left-right p -dimensional invariant subspaces of C . Moreover, Grassmannian versions of the Rayleigh quotient iteration are given for the generalized Hermitian eigenproblem, the Hamiltonian eigenproblem and the skew-Hamiltonian eigenproblem.

Keywords. Block Rayleigh quotient iteration, two-sided iteration, Grassmann manifold, generalized eigenproblem, Hamiltonian eigenproblem.

AMS subject classification. 65F15

1 Introduction

The Rayleigh quotient iteration (RQI) is a classical method for computing eigenvectors of a Hermitian matrix $A = A^H$ [Par74, Par98]. The RQI is a particular inverse iteration [Ips97] where the shift is the Rayleigh quotient evaluated at the current iterate. The Rayleigh quotient is an efficient shift because in a neighborhood of any eigenvector of A it yields a quadratic approximation of the corresponding eigenvalue. This remarkable property endows the iteration with cubic rate of convergence to the eigenvectors of A (see [Par74, Par98] or the sketch of proof in [AMSV02]). Thanks to its fast convergence, the RQI is particularly efficient for refining estimates of eigenvectors.

In some cases, one has to refine an estimate of a p -dimensional invariant subspace (or *eigenspace*) of A . A reason for considering an eigenspace instead of individual eigenvectors may be that the eigenvectors themselves are ill-conditioned while the subspace is not

*This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors. This work was supported by the US National Science Foundation under Grant OCI-0324944 and by the School of Computational Science of Florida State University through a postdoctoral fellowship.

[†]Department of Mathematical Engineering, Université catholique de Louvain, Avenue Georges Lemaître 4, B-1348 Louvain-la-Neuve, Belgium (<http://www.inma.ucl.ac.be/~absil,~vdooren>).

(see e.g. [Ste73]) or just because they are not relevant for the application. Several methods have been proposed for refining invariant subspace estimates. A quadratically convergent iteration for refining eigenspaces of arbitrary (possibly non-Hermitian) matrices was proposed by Chatelin [Cha84, Dem87]. It uses Newton's method for solving the Riccati equation obtained by expressing the eigenproblem in inhomogeneous coordinates. Similar Newton-like iterations for eigenspace refinement were obtained using a differential-geometric approach [EAS98, LE02, AMS04]; see [AMS07] for an overview. In [Smi97, AMSV02], it was shown that the RQI, originally defined on the set of one-dimensional subspaces of \mathbb{R}^n , can be generalized to operate on the set of p -dimensional subspaces of \mathbb{R}^n . The generalized iteration, called block-RQI or Grassmann-RQI (because the set of the p -dimensional subspaces of \mathbb{R}^n is termed a *Grassmann manifold*) converges locally cubically to the p -dimensional eigenspaces of $A = A^H$.

It is natural to ask whether the Grassmann-RQI method can be adapted to deal with non-Hermitian matrices. This is the topic of the present paper.

The underlying idea comes from Ostrowski's series of papers dedicated to the RQI [Ost59a]. Let C be a nonnormal matrix. Then the quadratic approximation property of the Rayleigh quotient is lost (and moreover the global convergence properties of the RQI become weaker, see [BS90]). This drawback was avoided by Ostrowski [Ost59b, Par74] by considering the bilateral Rayleigh quotient $\rho(y_L, y_R) := y_L^H C y_R / y_L^H y_R$ which displays the quadratic property in the neighborhood of the pairs of left-right nondefective eigenvectors of C . Using this Rayleigh quotient as a shift, he derived a two-sided iteration (see Algorithm 2.6 below) that operates on pairs of vectors (or pairs of one-dimensional subspaces, since the norm is irrelevant) and aims at converging to pairs of left-right eigenvectors of C . The rate of convergence is in cubic in nondegenerate cases. The possibility of solving the two-sided RQI equations approximately was investigated in [HS03].

In the present paper, we generalize Ostrowski's two-sided RQI to operate on pairs of p -dimensional subspaces (instead of one-dimensional subspaces in the original iteration). The new iteration, called *Two-Sided Grassmann-RQI (2sGRQI)*, converges locally cubically to the pairs of left-right p -dimensional eigenspaces of C (see Section 5). Comparison between Chatelin's iteration and the 2sGRQI (Section 6) shows that each method has its advantages and drawbacks. Main advantages of the 2sGRQI over Chatelin's iteration are the higher rate of convergence, the simultaneous computation of left and right eigenspaces, and the simpler structure of the Sylvester equations. On the other hand, the 2sGRQI does not behave satisfactorily when C is defective and it involves two Sylvester equations instead of one. We also show that in some structured eigenproblems, namely E -(skew-)Hermitian matrices with $E = \pm E^H$, a relation $\mathcal{Y}_L = E \mathcal{Y}_R$ between left and right subspaces is invariant by the 2sGRQI mapping (Section 7). In particular, this observation yields a modified one-sided Grassmann-RQI for the Hamiltonian eigenproblem. We report on numerical experiments in Section 8 and conclusions are drawn in Section 9.

2 Preliminaries

This paper uses a few elementary concepts related to the algebraic eigenvalue problem, such as principal vectors, Jordan blocks and nonlinear elementary divisors. A classical reference is [Wil65]. Notions of subspaces and distance between them can be found in [Ste73].

The superscript H denotes the conjugate transpose. In accordance with Parlett's conventions [Par74, Par98], we try to reserve the letter A for Hermitian matrices while C may denote any matrix. We use $\text{Grass}(p, n)$ to denote the Grassmann manifold of the p -dimensional subspaces of \mathbb{C}^n , \mathbb{P}^{n-1} to denote the projective space (i.e., the set of all one-dimensional subspaces of \mathbb{C}^n), and $\mathbb{C}_*^{n \times p}$ to denote the noncompact Stiefel manifold, i.e., the set of n -by- p matrices with full rank. The space spanned by the columns of $Y \in \mathbb{C}_*^{n \times p}$ is denoted by $[Y]$ and called the *span* of Y . The norm of a vector x is $\|x\| = \sqrt{x^H x}$. The spectral norm of a matrix T , denoted by $\|T\|$, is the largest singular value of T . The Hermitian angle $\angle(x, y)$ between two vectors x and y in \mathbb{C}^n is given by $\cos \angle(x, y) = \frac{|x^H y|}{\|x\| \|y\|}$ [Sch01]. The angle between a vector $y \in \mathbb{C}^n$ and a subspace \mathcal{X} spanned by $X \in \mathbb{C}_*^{n \times p}$ is $\angle(X, y) = \min_{x \in \mathcal{X}, \|x\|=1} \angle(x, y)$. The angle $\angle(X, Y)$ between two subspaces spanned by $X \in \mathbb{C}_*^{n \times p}$ and $Y \in \mathbb{C}_*^{n \times p}$ is defined as the largest principal angle between the two subspaces, given by $\cos \angle(X, Y) = \sigma_{\min}(\tilde{X}^H \tilde{Y})$ where \tilde{X} and \tilde{Y} are orthonormal bases for $[X]$ and $[Y]$, respectively, and σ_{\min} denotes the smallest singular value [GH06]. The following proposition is a generalization of [AMSV02, Th. 3.1] to the complex case.

Proposition 2.1 *Let $[X|X_\perp]$ be a unitary matrix of order n , with X of dimension $n \times p$, and let K be an $(n - p) \times p$ matrix. Then*

$$\tan \angle(X, X + X_\perp K) = \|K\|.$$

Proof. The matrix $\tilde{Y} = (X + X_\perp K)(I + K^H K)^{-1/2}$ is an orthonormal matrix with the same span as $X + X_\perp K$. It follows that $\cos \angle(X, X + X_\perp K) = \sigma_{\min}(X^H \tilde{Y}) = \sigma_{\min}(I + K^H K)^{-1/2} = (1 + \sigma_{\max}^2(K))^{-1/2} = (1 + \|K\|^2)^{-1/2}$. The conclusion follows from the trigonometric formula $\tan^2 a = (1 - \cos^2 a) / \cos^2 a$. \square

We now briefly recall some basic facts about invariant subspaces.

Definition 2.2 (eigenspaces) *Let \mathcal{X} be a p -dimensional subspace of \mathbb{C}^n and let $X = [X_1, X_2]$ be a unitary $n \times n$ matrix such that X_1 spans \mathcal{X} . Then $X^H C X$ may be partitioned in the form $X^H C X = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$ where $C_{11} \in \mathbb{C}^{p \times p}$. The subspace \mathcal{X} is an eigenspace (or invariant subspace) of C if $C_{21} = 0$, i.e., $C\mathcal{X} \subset \mathcal{X}$. By spectrum of \mathcal{X} , we mean the set of eigenvalues of C_{11} . We say that \mathcal{X} is a nondefective invariant subspace of C if C_{11} is nondefective. The invariant subspace \mathcal{X} is termed spectral if C_{11} and C_{22} have no eigenvalue in common [GLR86]. The eigenspaces of C^H are called left eigenspaces of C . We say that $(\mathcal{Y}_L, \mathcal{Y}_R)$ is a pair of spectral left-right eigenspaces of C if \mathcal{Y}_L and \mathcal{Y}_R are spectral left and right eigenspaces of C with the same spectrum.*

The span of $Y \in \mathbb{C}_*^{n \times p}$ is an eigenspace of C if and only if there exists a matrix M such that $CY = YM$. Each spectral eigenspace is *isolated*, i.e., there exists a ball in $\text{Grass}(p, n)$ centered on \mathcal{V} that does not contain any eigenspace of C other than \mathcal{V} . We will also need the following result [GV96, §7.6.3], of which we give an informative proof.

Proposition 2.3 *If $(\mathcal{Y}_L, \mathcal{Y}_R)$ is a pair of spectral left-right eigenspaces of C , then there exists an invertible matrix S such that the first p columns of S span \mathcal{Y}_R , the first p columns of S^{-H} span \mathcal{Y}_L , and $S^{-1}CS = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}$ with $D_1 \in \mathbb{C}^{p \times p}$.*

Proof. Let $(\mathcal{Y}_L, \mathcal{Y}_R)$ be a pair of spectral left-right eigenspaces of C . Then there exists X unitary such that $X^H C X = \begin{bmatrix} C_{11} & C_{12} \\ 0 & C_{22} \end{bmatrix}$, the first p columns of X span \mathcal{Y}_R , and C_{11} and C_{22} have no eigenvalue in common. Therefore, there exists a matrix L such that $C_{11}L - LC_{22} = -C_{12}$ [Ste73]. Let $S = X \begin{bmatrix} I & L \\ 0 & I \end{bmatrix}$. Then the first p columns of S span \mathcal{Y}_R . One easily checks that $S^{-1}CS = \begin{bmatrix} C_{11} & 0 \\ 0 & C_{22} \end{bmatrix}$. Moreover $C^H S^{-H} = S^{-H} \begin{bmatrix} C_{11}^H & 0 \\ 0 & C_{22}^H \end{bmatrix}$. Therefore the first p columns of S^{-H} span an eigenspace of C^H whose spectrum is the same as the one of \mathcal{Y}_R . That is, the first p columns of S^{-H} span \mathcal{Y}_L . \square

The Rayleigh quotient iteration (RQI) is a classical method for computing a single eigenvector of a Hermitian matrix A . It induces an iteration on the projective space \mathbb{P}^{n-1} that can be written as follows.

Algorithm 2.4 (RQI on projective space) *Let $A = A^H$ be an $n \times n$ matrix. Given \mathcal{S}_0 in the projective space \mathbb{P}^{n-1} , the RQI algorithm produces a sequence of elements of \mathbb{P}^{n-1} as follows. For $k = 0, 1, 2, \dots$,*

1. *Pick y in $\mathbb{C}^n \setminus \{0\}$ such that $\lfloor y \rfloor = \mathcal{S}_k$.*
2. *Compute the Rayleigh quotient $\rho_k = (y^H A y) / (y^H y)$.*
3. *If $A - \rho_k I$ is singular, then solve for its kernel and stop. Otherwise, solve the system*

$$(A - \rho_k I)z = y \tag{1}$$

for z .

4. *$\mathcal{S}_{k+1} := \lfloor z \rfloor$.*

It is shown in [BS89] that around each (isolated) eigenvector of A , there is a ball in which cubic convergence to the eigenvector is uniform. The size of the ball depends on the spacing between the eigenvalues. Globally, the RQI converges to an eigenvector for any initial point outside a certain set of measure zero described in [BS89].

The *Grassmann-Rayleigh Quotient Iteration* (GRQI) is a generalization of the RQI that operates on $\text{Grass}(p, n)$, the set of all p -dimensional subspaces of \mathbb{C}^n [AMSV02].

Algorithm 2.5 (GRQI) *Let $A = A^H$ be an $n \times n$ matrix. Given $\mathcal{Y}_0 \in \text{Grass}(p, n)$, the GRQI algorithm produces a sequence of p -dimensional subspaces of \mathbb{C}^n by iterating from \mathcal{Y}_0 the mapping $\text{Grass}(p, n) \rightarrow \text{Grass}(p, n) : \mathcal{Y} \mapsto \mathcal{Y}_+$ defined as follows.*

1. *Pick $Y \in \mathbb{C}_*^{n \times p}$ such that $\lfloor Y \rfloor = \mathcal{Y}$.*
2. *Solve the Sylvester equation*

$$AZ - Z(Y^H Y)^{-1} Y^H A Y = Y \tag{2}$$

for $Z \in \mathbb{C}^{n \times p}$.

3. *Define $\mathcal{Y}_+ := \lfloor Z \rfloor$.*

It is shown in [AMSV02] that the subspace \mathcal{Y}_+ does not depend on the choice of basis Y for \mathcal{Y} in the first step. This iteration converges cubically to the p -dimensional eigenspaces of A , which are the only fixed points.

When the matrix A is not normal, the stationary property of the Rayleigh quotient fails. Consequently, the convergence rate of the RQI can be at best quadratic. In order to recover

cubic convergence, Ostrowski [Ost59b] proposed a two-sided version of the RQI, formulated as follows in [Par74].

Algorithm 2.6 (Two-Sided RQI) *Let C be an n -by- n matrix. Pick initial vectors v_0 and u_0 satisfying $v_0^H u_0 \neq 0$, $\|v_0\| = \|u_0\| = 1$. For $k = 0, 1, 2, \dots$,*

1. *Compute $\rho_k = v_k^H C u_k / v_k^H u_k$.*
2. *If $C - \rho_k I$ is singular solve $y^H (C - \rho_k I) = 0$ and $(C - \rho_k I)x = 0$ for $y, x \neq 0$ and stop, otherwise*
3. *Solve both $v_{k+1}^H (C - \rho_k I) = v_k^H \nu_k$, $(C - \rho_k I)u_{k+1} = u_k \tau_k$, where ν_k and τ_k are normalizing factors.*
4. *If $v_{k+1}^H u_{k+1} = 0$, then stop and admit failure.*

The Two-Sided RQI converges with cubic rate to the pairs of left-right eigenvectors of C with linear elementary divisor [Par74].

3 Two-Sided GRQI

We propose the following generalization of the Two-Sided RQI, which we call the *Two-Sided Grassmann-Rayleigh Quotient Iteration (2sGRQI)*.

Algorithm 3.1 (2sGRQI) *Let C be an n -by- n matrix. Given $(\mathcal{Y}_{L_0}, \mathcal{Y}_{R_0}) \in \text{Grass}(p, n) \times \text{Grass}(p, n)$, the 2sGRQI algorithm produces a sequence of pairs of p -dimensional subspaces of \mathbb{C}^n by iterating from $(\mathcal{Y}_{L_0}, \mathcal{Y}_{R_0})$ the mapping $(\mathcal{Y}_L, \mathcal{Y}_R) \mapsto (\mathcal{Y}_{L+}, \mathcal{Y}_{R+})$ defined as follows.*

1. *Pick Y_L and Y_R in $\mathbb{C}_*^{n \times p}$ such that $[Y_L] = \mathcal{Y}_L$ and $[Y_R] = \mathcal{Y}_R$.*
2. *Solve the Sylvester equations*

$$C Z_R - \underbrace{Z_R (Y_L^H Y_R)^{-1} Y_L^H C Y_R}_{R_R} = Y_R \quad (3a)$$

$$Z_L^H C - \underbrace{Y_L^H C Y_R (Y_L^H Y_R)^{-1} Z_L^H}_{R_L} = Y_L^H \quad (3b)$$

for Z_L and Z_R in $\mathbb{C}^{n \times p}$.

3. Define $\mathcal{Y}_{L+} := [Z_L]$ and $\mathcal{Y}_{R+} := [Z_R]$.

In point 1, one has to choose bases for \mathcal{Y}_L and \mathcal{Y}_R . There are infinitely many possibilities. Indeed, if Y is a basis of \mathcal{Y} , then $\{Y M : M \in \mathbb{C}_*^{p \times p}\}$ is the (infinite) set of all bases of \mathcal{Y} . Therefore, one has to make sure that \mathcal{Y}_{L+} and \mathcal{Y}_{R+} do not depend on the choice of basis. By a straightforward adaptation of the development carried out in [AMSV02] for the GRQI algorithm, if (Y_L, Y_R, Z_L, Z_R) solve (3) then $(Y_L M, Y_R N, Z_L M, Z_R N)$ also solve (3) for all M, N in $\mathbb{C}_*^{p \times p}$. Hence, the spans of Z_L and Z_R only depend on \mathcal{Y}_L and \mathcal{Y}_R , and not on the choice of the bases Y_L and Y_R .

In point 2, the matrix $Y_L^H Y_R$ may not be invertible. This corresponds to point 4 in the Two-Sided RQI (Algorithm 2.6). However, if $(\mathcal{Y}_L, \mathcal{Y}_R)$ is a pair of spectral left-right eigenspaces of C , then $Y_L^H Y_R$ is invertible (as a consequence of Proposition 2.3), and by continuity invertibility holds on a neighborhood of the pair of eigenspaces.

In point 2, the (uncoupled) Sylvester equations (3) may fail to admit one and only one solution. This situation happens if and only if (Y_R, Y_L) belongs to the set

$$\begin{aligned}\mathcal{S} &:= \{(Y_L, Y_R) \in \mathbb{C}_*^{n \times p} \times \mathbb{C}_*^{n \times p} : R_R \text{ exists and } \sigma(C) \cap \sigma(R_R) \neq \emptyset\} \\ &= \bigcup_{\lambda \in \sigma(C)} \{(Y_L, Y_R) \in \mathbb{C}_*^{n \times p} \times \mathbb{C}_*^{n \times p} : R_R \text{ exists and } \det(R_R - \lambda I) = 0\};\end{aligned}$$

this follows directly from the characterization of the eigenvalues of Sylvester operators [Ste73, Th. 4.4]. Since \mathcal{S} is the finite union of algebraic sets, it has measure zero and the interior of its closure is empty. This means that if (\hat{Y}_L, \hat{Y}_R) does not yield a unique solution, then there exists, arbitrarily close to (\hat{Y}_L, \hat{Y}_R) , a pair (Y_L, Y_R) and a neighborhood of this pair on which the solution (Z_L, Z_R) of (3) exists and is unique. In our numerical experiments, when such a singularity occurs (i.e., when the solution of the Sylvester equations returned by Matlab contains **Inf**'s or **NaN**'s), we slightly perturb the system. A justification for this technique is given in [AMSV02] and the numerical tests performed in Section 8 illustrate that the technique works well in practice.

In point 3, if Z_L or Z_R is not full rank, then $(\mathcal{Y}_{L+}, \mathcal{Y}_{R+})$ does not belong to $\text{Grass}(p, n) \times \text{Grass}(p, n)$. A tall $n \times p$ matrix Z is rank deficient if and only if all its $p \times p$ minors are zero. Therefore, the set

$$\mathcal{D} := \{(Y_L, Y_R) : \text{rank}(Z_L) < p \text{ or } \text{rank}(Z_R) < p\}$$

is a subset of a finite union of algebraic sets. So here again, Z_L and Z_R are full rank for a generic choice of Y_L, Y_R .

In practice, only a few iterates will be computed. In finite precision arithmetic, the iterates no longer improve after a few (typically two or three) iterations because of numerical errors (see numerical experiments in Section 8). Stopping criteria can rely on the principal angles between two successive iterates and on the principal angles between \mathcal{Y}_R and $A\mathcal{Y}_R$ or \mathcal{Y}_L and $A^H\mathcal{Y}_L$.

4 Practical implementation

The major computational task in both GRQI (Algorithm 2.5) and 2sGRQI (Algorithm 3.1) is to solve the Sylvester equations. For GRQI, it is recommended to choose an orthonormal basis Y (i.e., $Y^H Y = I_p$) that makes $Y^H A Y$ diagonal. This requires solving a p -dimensional eigenproblem, which is cheap when p is small. With $Y^H Y = I$ and $Y^H A Y$ diagonal, the GRQI equation (2) decouples into p linear systems for the p columns of Z . We refer to [AMSV02] for details.

The case of 2sGRQI (Algorithm 3.1) is quite different. The matrices R_R and R_L in the 2sGRQI equations (3) are not Hermitian and they may not be diagonalizable. A possible approach to solving an equation such as (3a) is to reduce it to a certain triangular structure by means of unitary transformations and solve the new system of equations using back substitution, as described in [GLAM92]. However, we observed in numerical experiments that this technique tends to yield rather inaccurate results when the iterates get close to a solution (the final error was sometimes around 10^{-11} whereas the machine epsilon was approximately $2.2 \cdot 10^{-16}$, to be compared with the results in Table 1). The reason seems to lie in the fact

that the norm of the solution z to the equation $(C - \rho I)z = x$ becomes very sensitive to ρ when ρ gets close to an eigenvalue of C . The magic of the classical RQI is that the direction of z is well-conditioned, as pointed out by Peters and Wilkinson [PW79]. However, in the block case, the magic weakens because, in view of the workings of back substitution, the large numerical error in the norm of any one column of Z jeopardizes the accuracy of the other columns whose computation depends on that column.

Consequently, we recommend reducing the small block shifts R_R and R_L in the 2sGRQI equations (3) to quasi-diagonal form, or to (complex) diagonal form if complex arithmetic is available. To fix ideas, assume complex arithmetic and consider the first equation, (3a), namely

$$CZ_R - Z_R R_R = Y_R.$$

Assuming that R_R is nondefective, let $R_R = W_R \text{diag}(\rho_1, \dots, \rho_p) W_R^{-1}$ be an eigenvalue decomposition of R_R . Multiplying (3a) on the right by W_R yields

$$C\tilde{Z}_R - \tilde{Z}_R \text{diag}(\rho_1, \dots, \rho_p) = \tilde{Y}_R$$

where $\tilde{Y}_R = Y_R W_R$ and $\tilde{Z}_R = Z_R W_R$. The advantage of this reformulation of (3a) is that it yields p decoupled equations

$$(C - \rho_i I)\tilde{Z}_R e_i = \tilde{Y}_R e_i$$

for each column $\tilde{Z}_R e_i$ of \tilde{Z}_R . Back propagation is thus no longer needed. A drawback is that this technique does not work when R_R is defective, and numerical errors on W_R may become large when R_R is close to being defective. Nevertheless, in our extensive numerical experiments on randomly chosen matrices, these difficulties were not noticed (see Section 8).

The same kind of discussion applies to the left equation (3b). Note that since $R_L = (Y_L^H Y_R) R_R (Y_L^H Y_R)^{-1}$ is a similarity transformation of R_R , we have that $W_L = (Y_L^H Y_R) W_R$ is a matrix of eigenvector of R_L . Hence, the eigendecomposition of R_L is readily obtained from that of R_R .

5 Local convergence

The following local convergence analysis can be thought of as a two-sided generalization of the proof of cubic convergence of the block-RQI (equivalent to the Grassmann-RQI of [AMSV02]) given in [Smi97].

Let $(\mathcal{V}_L, \mathcal{V}_R)$ be a pair of spectral left-right eigenspaces of C , and let V_L and V_R be corresponding eigenbases. We assume that the eigenspaces are nondefective, that is, the matrix $(V_L^H V_R)^{-1} V_L^H C V_R$ is diagonalizable by a similarity transformation. Since $(\mathcal{V}_L, \mathcal{V}_R)$ is nondefective, it follows that for all \mathcal{Y}_L and \mathcal{Y}_R sufficiently close to \mathcal{V}_L and \mathcal{V}_R , the block Rayleigh quotients R_R and R_L are diagonalizable by similarity transformations W_R and W_L . Equations (3) thus can be solved in two steps: (i) diagonalize the small block Rayleigh quotients, hence decoupling the equations and reducing them to classical two-sided RQI equations; (ii) solve the decoupled two-sided RQI equations, yielding matrices $Z_L W_L$ and $Z_R W_R$ that span \mathcal{Y}_{L+} and \mathcal{Y}_{R+} . The key of the convergence analysis is an “oblique” generalization of [Ste01, Th. 2], showing that the angles between the right Ritz vectors (the columns of $Y_R W_R$) and the “corresponding” right eigenvectors of A are of the order of the largest principal angle between \mathcal{Y}_R and \mathcal{V}_R , and likewise for the left Ritz vectors and eigenvectors; see Lemma 5.1.

Then the result follows quite directly from the cubic convergence of the non-block two-sided RQI.

Lemma 5.1 *Let (λ, x) be an eigenpair of an $n \times n$ matrix C . Let Y_R and Y_L be orthonormal $n \times p$ matrices, $p < n$, such that $Y_L^H Y_R$ is invertible. Let w_R be an eigenvector of*

$$B := (Y_L^H Y_R)^{-1} Y_L^H C Y_R$$

associated with the eigenvalue of B that is closest to λ . Then

$$\sin \angle(Y_R w_R, x) \leq \left[1 + \frac{2(\cos \delta)^{-1} r_L \alpha_\delta(\epsilon)}{\text{sep}(w_R^H B w_R, w_{R\perp}^H B w_{R\perp}) - r_L \gamma_\delta(\epsilon)} \right] (1 + \tan \delta) \epsilon$$

where $\epsilon := \sin \angle(Y_R, x)$ is the angle between the direction of x and the span of Y_R , $\delta := \angle(Y_R, Y_L)$ is the largest principal angle between the spans of Y_R and Y_L , $\alpha_\delta(\epsilon) := \frac{1}{\sqrt{1-\epsilon^2}-\epsilon \tan \delta}$ satisfies $\lim_{\epsilon \rightarrow 0} \alpha_\delta(\epsilon) = 1$, $\gamma_\delta(\epsilon) := (\cos \delta (\sqrt{1-\epsilon^2} - \epsilon \tan \delta))^{-1} (1 + \tan \delta) \epsilon$ satisfies $\lim_{\epsilon \rightarrow 0} \gamma_\delta(\epsilon) = 0$, and $r_L := \|Y_{L\perp}^H A^H Y_L\|$ where $Y_{L\perp} \in \mathbb{C}^{n \times (n-p)}$ is an orthonormal basis of the orthogonal complement of the span of Y_L .

Proof. It is readily checked that the statement is not affected by a unitary change of coordinates in \mathbb{C}^n . Therefore, without loss of generality, we work in a unitary coordinate system such that $Y_R = \begin{bmatrix} I_p \\ 0_{(n-p) \times p} \end{bmatrix}$. Let $Y_{L\perp} \in \mathbb{C}^{n \times (n-p)}$ and $Y_{R\perp} \in \mathbb{C}^{n \times (n-p)}$ be orthonormal bases of the orthogonal complements of the spans of Y_L and Y_R , respectively. Assume without loss of generality that the eigenvector x has unit norm. Consider the block decompositions $x = \begin{bmatrix} x_a \\ x_b \end{bmatrix}$ and $Y_L = \begin{bmatrix} Y_{La} \\ Y_{Lb} \end{bmatrix}$. Consider also the decomposition $x = Y_R x_R + Y_{L\perp} x_{L\perp}$, which yields

$$x_R := (Y_L^H Y_R)^{-1} Y_L^H x, \quad x_{L\perp} := (Y_{R\perp}^H Y_{L\perp})^{-1} Y_{R\perp}^H x.$$

Since $\epsilon = \sin \angle(Y_R, x)$, we have $\|x_a\|^2 = 1 - \epsilon^2$ and $\|x_b\| = \epsilon$. We also have $(Y_L^H Y_R)^{-1} Y_L^H = \begin{bmatrix} I & T \end{bmatrix}$ where $T = (Y_{La})^{-1} Y_{Lb}$. It follows from Proposition 2.1 that $\|T\| = \tan \delta$. We also obtain

$$Y_R x_R = \begin{bmatrix} I \\ 0 \end{bmatrix} \begin{bmatrix} I & T \end{bmatrix} x = \begin{bmatrix} x_a + T x_b \\ 0 \end{bmatrix}.$$

Acceptable choices for $Y_{L\perp}$ and $Y_{R\perp}$ are $Y_{L\perp} = \begin{bmatrix} -T \\ I_{n-p} \end{bmatrix} (I_{n-p} + T^H T)^{-1/2}$ and $Y_{R\perp} = \begin{bmatrix} 0_{p \times (n-p)} \\ I_{n-p} \end{bmatrix}$. This yields $x_{L\perp} = (I_{n-p} + T^H T)^{1/2} x_b$ and thus $\|x_L\| \leq \sqrt{1 + \tan^2 \delta} \epsilon$.

Since $\sin \angle(u, v) \leq \sin \angle(u, w) + \sin \angle(w, v)$ for all $u, v, w \in \mathbb{C}_0^n$, we have

$$\angle(Y_R w_R, x) \leq \angle(Y_R w_R, Y_R x_R) + \angle(Y_R x_R, x). \quad (4)$$

Let us first consider the second term in (4). Since

$$\sin \angle(Y_R x_R, x) \leq \|Y_R x_R - x\| \leq \|Y_R x_R - \begin{bmatrix} x_a \\ 0 \end{bmatrix}\| + \left\| \begin{bmatrix} x_a \\ 0 \end{bmatrix} - x \right\|,$$

it follows that

$$\sin \angle(Y_R x_R, x) \leq \|Tx_b\| + \|x_b\| \leq \tan \delta \epsilon + \epsilon = (1 + \tan \delta) \epsilon. \quad (5)$$

Note also for later use that, for all small ϵ such that $\sqrt{1 - \epsilon^2} > \epsilon \tan \delta$, we also obtain that $\|x_R\| \geq \|x_a\| - \|Tx_b\| \geq \sqrt{1 - \epsilon^2} - \epsilon \tan \delta$.

We now tackle the first term in (4). Since Y_R is orthonormal, it follows that $\angle(Y_R w_R, Y_R x_R) = \angle(w_R, x_R)$. Pre-multiplying the equation $Cx = \lambda x$ by $(Y_L^H Y_R)^{-1} Y_L^H$ yields

$$(Y_L^H Y_R)^{-1} Y_L^H C(Y_R x_R + Y_{L\perp} x_{L\perp}) = x_R \lambda,$$

which can be rewritten as

$$(B + E) \hat{x}_R = \lambda \hat{x}_R,$$

where $\hat{x}_R := x_R \|x_R\|^{-1}$ and

$$E := (Y_L^H Y_R)^{-1} Y_L^H A Y_{L\perp} x_{L\perp} \|x_R\|^{-1} \hat{x}_R^H.$$

Then, by [JS00, Th. 5.1],

$$\sin \angle(w_R, \hat{x}_R) \leq \tan \angle(w_R, \hat{x}_R) \leq \frac{2\|E\|}{\text{sep}(w_R^H B w_R, (w_R)_\perp^H B (w_R)_\perp) - 2\|E\|}$$

if the bound is smaller than 1. The expression of the bound can be simplified using

$$\begin{aligned} \|E\| &= \|(Y_L^H Y_R)^{-1} Y_L^H C Y_{L\perp} x_{L\perp}\| \|x_R\|^{-1} \leq \|(Y_L^H Y_R)^{-1}\| \|Y_{L\perp}^H C^H Y_L\| \|x_{L\perp}\| \|x_R\|^{-1} \\ &\leq \frac{1}{\cos \delta} r_L (1 + \tan \delta) \epsilon \frac{1}{\sqrt{1 - \epsilon^2} - \epsilon \tan \delta}, \end{aligned}$$

where we have used the bound $\sqrt{1 + \tan^2 \delta} \leq (1 + \tan \delta)$ that holds for all $\delta \in [0, \frac{\pi}{2})$. Replacing all these results in (4) yields the desired bound. \square

Theorem 5.2 *Let $(\mathcal{V}_L, \mathcal{V}_R)$ be a pair of p -dimensional spectral nondefective left-right eigenspaces of an $n \times n$ matrix C (Definition 2.2). Then there is a neighborhood \mathcal{N} of $(\mathcal{V}_L, \mathcal{V}_R)$ in $\text{Grass}(p, n) \times \text{Grass}(p, n)$ and a $c > 0$ such that, for all $(\mathcal{Y}_L, \mathcal{Y}_R) \in \mathcal{N}$, the subspaces \mathcal{Y}_{L+} and \mathcal{Y}_{R+} produced by the 2sGRQI mapping (Algorithm 3.1) satisfy*

$$\angle(\mathcal{Y}_{L+}, \mathcal{V}_L) + \angle(\mathcal{Y}_{R+}, \mathcal{V}_R) \leq c(\angle(\mathcal{Y}_L, \mathcal{V}_L) + \angle(\mathcal{Y}_R, \mathcal{V}_R))^3.$$

Proof. Since the pair of eigenspaces is assumed to be spectral, it follows that $\angle(V_L, V_R) < \pi/2$. Therefore, taking the neighborhood \mathcal{N} sufficiently small, one has $\angle(Y_R, Y_L) \leq \delta' < \pi/2$. Moreover, since the pair of eigenspaces is assumed to be nondefective, it follows that the eigenbases V_R and V_L have full rank. Note that for each column x of V_R , we have $\angle(Y_R, x) \leq \angle(Y_R, V_R)$. Lemma 5.1 implies that for any $c_1 > 1 + \tan \delta'$, there exists an $\epsilon > 0$ such that, for all $(\mathcal{Y}_L, \mathcal{Y}_R)$ with $\angle(\mathcal{Y}_L, \mathcal{V}_L) + \angle(\mathcal{Y}_R, \mathcal{V}_R) < \epsilon$, the angle $\angle(Y_R w_R, x)$ between x and the nearest Ritz vector $Y_R w_R$ satisfies $\angle(Y_R w_R, x) \leq c_1 \angle(Y_R, x) \leq c_1 \angle(Y_R, V_R)$. Next, represent the subspaces \mathcal{Y}_L and \mathcal{Y}_R by their Ritz vectors, which decouples (3a) into p two-sided RQI equations. By taking ϵ sufficiently small, it follows from the cubic convergence of the two-sided RQI that there exists $c_2 > 0$ such that, for each column x of V_R , we have $\angle((z_R)_i, x) < c_2 (c_1 \angle(Y_R, V_R))^3$ for at least one column $(z_R)_i$ of Z_R . It follows that $\angle(Z_R, V_R) < c_3 c_2 (c_1 \angle(Y_R, V_R))^3$ where c_3 is a constant that depends on the conditioning of the basis V_R . A similar reasoning applies to the left subspace. \square

6 Comparisons with Newton-based approaches

It has been long known (see, e.g., Peters and Wilkinson [PW79]) that the RQI can be viewed as a Newton iteration. In fact, the RQI *is* a Newton method in a certain differential-geometric sense [AMS07]. However, the strict interpretation of RQI as a Newton method disappears in the block case, as pointed out in [AMSV02], so much so that the Grassmann-RQI can be considered as distinct from the Newton approach.

Several Newton-based approaches for the general (non-Hermitian) eigenvalue problem have been proposed in the literature. In particular, the well-known Jacobi-Davidson approach can be viewed as a Newton method within a sequential subspace algorithm; see, e.g., [LE02, AMS07]. Here we discuss specifically the Newton method proposed by Chatelin [Cha84] for refining eigenspace estimates. The reasoning can be explained as follows. An $n \times p$ matrix Y spans an eigenspace of C if and only if there exists a $p \times p$ matrix M such that

$$CY = YM. \quad (6)$$

However, any subspace admits infinitely many bases, and the solutions Y of (6) are thus not isolated. A way to remove the freedom in the choice of basis is to impose on Y a normalization condition $W^H Y = I$ where W is a given full-rank $n \times p$ matrix. Then (6) becomes

$$F(Y) := CY - Y(W^H CY) = 0 \quad (7)$$

where the unknown Y is normalized by $W^H Y = I$. The Newton iteration for solving (7) is given by

$$(I - YW^H)C\Delta - \Delta(W^H CY) = -F(Y), \quad W^H \Delta = 0 \quad (8)$$

$$Y_+ := Y + \Delta. \quad (9)$$

If the basis Y is chosen orthonormal and $W := Y$, then (8) becomes

$$\Pi C \Pi \Delta - \Delta(Y^H CY) = -\Pi CY, \quad Y^H \Delta = 0 \quad (10)$$

where $\Pi := I - YY^H$. The resulting algorithm admits an interpretation as a Newton method on the Grassmann manifold [AMS07]. The rate of convergence is quadratic in general (cubic when C is Hermitian).

The constraint $Y^H \Delta = 0$ can be addressed by setting $\Delta = Y_\perp K$, where Y_\perp is an orthonormal matrix with $Y^H Y_\perp = 0$ and K is an $(n-p) \times p$ matrix; see, e.g., [Dem87]. Then $Y^H \Delta = 0$ is trivially satisfied and equation (10) becomes

$$(Y_\perp^H CY_\perp)K - K(Y^H CY) = -Y_\perp^H CY, \quad (11)$$

i.e., a Sylvester equation without constraints on the unknown K . As pointed out in [AMSV02], solving (11) takes $O(n^3)$ operations even when C is condensed (e.g. tridiagonal) because $Y_\perp^H CY_\perp$ is a large dense $(n-p) \times (n-p)$ matrix. However, Lundström and Eldén proposed an algorithm [LE02, alg. 2] for solving (10) that does not require the computation of $Y_\perp^H CY_\perp$. It takes $O(np^2)$ operations to solve (10) when C is block diagonal of sufficiently moderate block size and $O(n^2p)$ when C is Hessenberg. The complexity of the 2sGRQI method (Algorithm 3.1) is of the same order.

A theoretical comparison between algorithms based on inverse iteration and on Newton does not reveal that one approach has a clear edge over the other. Among the advantages of the 2sGRQI method (Algorithm 3.1) over Chatelin's method, one can mention that the convergence of 2sGRQI is cubic instead of quadratic, and that a pair of left-right eigenspaces is computed instead of just a right-eigenspace. On the other hand, Chatelin's method admits a convergence analysis when the target eigenspace is defective [Dem87, AMS04], and it requires solving only one Sylvester equation instead of two in 2sGRQI. However, we show in Section 7 that one Sylvester equation suffices for 2sGRQI on some important structured eigenproblems.

7 Structured eigenproblems

In this section, we show that the 2sGRQI induces particular one-sided formulations for some structured eigenproblems.

7.1 E -Hermitian eigenproblem

Let C be an $n \times n$ matrix. If there exists an invertible matrix E such that

$$EC = C^H E, \quad (12)$$

then we say that C is E -Hermitian. If C is E -Hermitian, then its left and right eigenspaces are related by the action of E . Indeed, let S be a (complex) matrix of principal vectors of C , i.e.,

$$CS = SD$$

where D is a (complex) Jordan matrix; then, from (12), one obtains $C^H(ES) = (ES)D$.

The case where E is Hermitian or skew-Hermitian, i.e., $E^H = \pm E$, is of particular interest because, as we show in the next proposition, the relation $\mathcal{Y}_L = E\mathcal{Y}_R$ is invariant under the 2sGRQI (Algorithm 3.1). Therefore, if $\mathcal{Y}_L = E\mathcal{Y}_R$, it is not necessary to solve both (3a) and (3b): just solve (3a) to get \mathcal{Y}_{R+} , and obtain \mathcal{Y}_{L+} as $\mathcal{Y}_{L+} := E\mathcal{Y}_{R+}$. Moreover, since the pairs of left-right eigenspaces of C also satisfy $\mathcal{V}_L = E\mathcal{V}_R$, Theorem 5.2 also applies.

Proposition 7.1 *Let E be invertible with $E^H = \pm E$ and let C be E -Hermitian, i.e., $EC = C^H E$. If $Y_L = EY$, $Y_R = Y$, and Z satisfies*

$$\boxed{CZ - Z(Y^H EY)^{-1}(Y^H E C Y) = Y}, \quad (13)$$

then $Z_L = EZ$ and $Z_R = Z$ satisfy the 2sGRQI equations (3). Hence, if $\mathcal{Y}_L = E\mathcal{Y}_R$, then $\mathcal{Y}_{L+} = E\mathcal{Y}_{R+}$. Moreover, the subspace iteration $[Y] \mapsto [Z]$ defined by (13) converges locally cubically to the spectral nondefective right-eigenspaces of C .

Proof. It is easy to check that replacing $Y_R := Y$, $Z_R := Z$, $Y_L := EY_R$, $Z_L := EZ_R$ in (3a) and (3b) yields (13) in both cases. In order to prove cubic convergence, it is sufficient to notice that the pairs $(\mathcal{V}_L, \mathcal{V}_R)$ of eigenspaces satisfy $\mathcal{V}_L = E\mathcal{V}_R$, as was shown above. Therefore, if \mathcal{Y} is close to \mathcal{V}_R , then the pair $(\mathcal{Y}_L, \mathcal{Y}_R) := (E\mathcal{Y}, \mathcal{Y})$ is close to $(\mathcal{V}_L, \mathcal{V}_R)$ and local cubic convergence to \mathcal{V}_R follows from Theorem 5.2. \square

The discussion in Section 4 on solving Sylvester equations applies likewise to (13).

Generalized Hermitian eigenproblem

Using Proposition 7.1, we show that the 2sGRQI yields a Grassmannian RQI for the Hermitian generalized eigenproblem $A\mathcal{V} \subset B\mathcal{V}$ which does not involve an explicit computation of $B^{-1}A$. Let A and B be two Hermitian n -by- n matrices with B invertible. Consider the problem of finding a p -dimensional subspace \mathcal{V} such that $A\mathcal{V} \subset B\mathcal{V}$. Let $V \in \mathbb{C}^{n \times p}$ be a basis for \mathcal{V} , then $A\mathcal{V} \subset B\mathcal{V}$ if and only if there is a matrix M such that $AV = BVM$. Equivalently, V spans a right-eigenspace of $B^{-1}A$, i.e.,

$$B^{-1}AV = VM.$$

The problem is thus to find a right-eigenspace of $C := B^{-1}A$. The conditions in Proposition 7.1 are satisfied with $E := B$. The modified GRQI equation (13) becomes

$$\boxed{AZ - BZ(Y^H BY)^{-1}(Y^H AY) = BY} \quad (14)$$

and the subspace iteration $[Y] \mapsto [Z]$ converges locally cubically to the spectral nondefective eigenspaces of $B^{-1}A$. In particular, $B^{-1}A$ is nondefective when A or B is positive definite.

Skew-Hamiltonian eigenproblem

Let T be a skew-Hamiltonian matrix, i.e., $(TJ)^H = -TJ$, where $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$, see e.g. [BBMX02]. Equivalently, $JT = T^H J$, i.e., T is J -Hermitian. Conditions in Proposition 7.1 are satisfied with $C := T$ and $E := J$. The modified GRQI equation (13) becomes

$$\boxed{TZ - Z(Y^H JY)^{-1}(Y^H JTY) = Y} \quad (15)$$

and the subspace iteration $[Y] \mapsto [Z]$ converges locally cubically to the spectral nondefective right-eigenspaces of T .

7.2 E -skew-Hermitian eigenproblem

Let E be an invertible $n \times n$ matrix and let C be an E -skew-Hermitian $n \times n$ matrix, namely

$$EC = -C^H E. \quad (16)$$

We saw in the previous section that the corresponding left and right eigenspaces of E -Hermitian matrices are related by a multiplication by E . The case of E -skew-Hermitian matrices is slightly different.

Proposition 7.2 *Let C be an E -skew-Hermitian matrix. Then the spectrum of C is symmetric with respect to the imaginary axis. In other words, if λ is an eigenvalue of C , then so is $-\bar{\lambda}$. Moreover, if \mathcal{V}_L and \mathcal{V}_R are left and right eigenspaces of C whose spectra are the symmetric image one of the other with respect to the imaginary axis, then $\mathcal{V}_L = E\mathcal{V}_R$.*

Proof. Letting S be an invertible matrix of principal vectors of C , i.e.,

$$CS = SD \quad (17)$$

where D is a Jordan matrix, (16) yields

$$C^H ES = ES(-D). \quad (18)$$

Hence, the matrix $-D$ is a Jordan matrix of C^H . Therefore, if λ is an eigenvalue of C , then $-\lambda$ is an eigenvalue of C^H , and thus $-\bar{\lambda}$ is an eigenvalue of C . Moreover, equations (17) and (18) show that if \mathcal{V} is a right-eigenspace of C with eigenvalues $\lambda_{i_1}, \dots, \lambda_{i_p}$, then $E\mathcal{V}$ is a left-eigenspace of C with eigenvalues $-\bar{\lambda}_{i_1}, \dots, -\bar{\lambda}_{i_p}$. \square

Consequently, letting \mathcal{V} be a spectral right-eigenspace of C , we have that $(E\mathcal{V}, \mathcal{V})$ forms a pair of spectral left-right eigenspaces of C if and only if the spectrum of \mathcal{V} is symmetric with respect to the imaginary axis. We call such an invariant subspace \mathcal{V} a *full eigenspace* of the E -skew-Hermitian matrix C .

If E is Hermitian or skew-Hermitian, then the relation $\mathcal{Y}_L = E\mathcal{Y}_R$ is invariant by the 2sGRQI (Algorithm 3.1), as we show in the forthcoming proposition. Therefore, if $\mathcal{Y}_L = E\mathcal{Y}_R$, it is sufficient to solve (3a) only, and then compute $\mathcal{Y}_{L+} := E\mathcal{Y}_{R+}$. Moreover, the 2sGRQI iteration restricted to the pairs $(\mathcal{Y}_L, \mathcal{Y}_R) = (E\mathcal{Y}, \mathcal{Y})$ converges locally cubically to the full nondefective eigenspaces of C .

Proposition 7.3 *Let E be invertible with $E^H = \pm E$ and let C be E -skew-Hermitian, i.e., $EC = -C^H E$. If $Y_L = EY$ and $Y_R = Y$, then $Z_L = -EZ$ and $Z_R = Z$ satisfy the 2sGRQI equations (3) with*

$$\boxed{CZ - Z(Y^H EY)^{-1}(Y^H ECY) = Y.} \quad (19)$$

Therefore, if $\mathcal{Y}_L = E\mathcal{Y}_R$, then $\mathcal{Y}_{L+} = E\mathcal{Y}_{R+}$.

Moreover, let \mathcal{V} be a full nondefective right-eigenspace of C (which means that the eigenvalues of $C|_{\mathcal{V}}$ have the same multiplicity as in C , the spectrum of $C|_{\mathcal{V}}$ is symmetric with respect to the imaginary axis, and $C|_{\mathcal{V}}$ is nondefective). Then the subspace iteration $[Y] \mapsto [Z]$ defined by (13) converges locally cubically to \mathcal{V} .

Note that this proposition differs from Proposition 7.1 in two points: $Z_L = -EZ$ and the specification that \mathcal{V} must be full.

Proof. It is easy to check that replacing $Y_R := Y$, $Z_R := Z$, $Y_L := EY_R$, $Z_L := -EZ_R$ in (3a) and (3b) yields (19) in both cases. In order to prove cubic convergence, it is sufficient to notice that the pairs $(\mathcal{V}_L, \mathcal{V}_R)$ of full nondefective left-right eigenspaces satisfy $\mathcal{V}_L = E\mathcal{V}_R$, as was shown above. Therefore, if \mathcal{Y} is close to \mathcal{V}_R , then the pair $(\mathcal{Y}_L, \mathcal{Y}_R) := (E\mathcal{Y}, \mathcal{Y})$ is close to $(\mathcal{V}_L, \mathcal{V}_R)$ and local cubic convergence to \mathcal{V} follows from Theorem 5.2. \square

Skew-Hermitian eigenproblem

Let Ω be skew-Hermitian. Then we have $EC = -C^H E$ with $C := \Omega$ and $E := I$. The modified GRQI equation (19) becomes

$$\boxed{\Omega Z - Z(Y^H Y)^{-1}(Y^H \Omega Y) = Y.} \quad (20)$$

This is simply the classical GRQI equation (2). This is not surprising as skew-Hermitian matrices are normal matrices.

Hamiltonian eigenproblem

Let H be Hamiltonian, i.e., $(HJ)^H = HJ$, where $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$. This is equivalent to $JH = -H^H J$. Thus we have $EC = -C^H E$ with $C := H$ and $E := J$, and the modified GRQI equation (19) reads

$$\boxed{HZ - Z(Y^H JY)^{-1}(Y^H JHY) = Y.} \quad (21)$$

Proposition 7.3 implies that the subspace iteration with iteration mapping $[Y] \mapsto [Z]$ defined by (21) converges locally cubically to the full nondefective right-eigenspaces of H .

7.3 The generalized eigenvalue problem

We briefly discuss the application of the 2sGRQI concept to the generalized eigenvalue problem. Let $A, B \in \mathbb{C}^{n \times n}$. The generalized eigenvalue problem consists in finding the nontrivial solutions of the equation $Ax = \lambda Bx$. Corresponding to the notion of invariant subspace for a single matrix, we have the notion of a *deflating subspace*, see e.g. [Ste73, GV96]. The p -dimensional subspace \mathcal{X} is deflating for the pencil $A - \lambda B$ if there exists a p -dimensional subspace \mathcal{Y} such that

$$A\mathcal{X}, B\mathcal{X} \subset \mathcal{Y}. \quad (22)$$

Here we suppose that the pencil $A - \lambda B$ is nondegenerate, i.e., $\det(A - \lambda B)$ is not trivially zero. Then there exists α and β such that $\hat{B} := \alpha B - \beta A$ is invertible. Now take γ, δ such that $\alpha\delta - \gamma\beta \neq 0$ and let $\hat{A} := \gamma B - \delta A$. Then (22) is equivalent to

$$\begin{aligned} \hat{B}^{-1}\hat{A}\mathcal{X} &\subset \mathcal{X} \\ \hat{B}\mathcal{X} &= \mathcal{Y}, \end{aligned}$$

i.e., \mathcal{X} is an invariant subspace of $\hat{B}^{-1}\hat{A}$. Replacing this expression for C in (3), one obtains after some manipulations

$$\hat{A}Z_R\hat{Y}_L^H\hat{B}Y_R - \hat{B}Z_R\hat{Y}_L^H\hat{A}Y_R = \hat{B}Y_R \quad (23a)$$

$$\hat{A}^H\hat{Z}_LY_R^H\hat{B}^H\hat{Y}_L - \hat{B}^H\hat{Z}_LY_R^H\hat{A}^H\hat{Y}_L = \hat{B}^H\hat{Y}_L \quad (23b)$$

where $\hat{Y}_L := \hat{B}^{-H}Y_L$ and $\hat{Z}_L := \hat{B}^{-H}Z_L$. It yields an iteration for which Y_R and \hat{Y}_L locally cubically converge to pairs of left-right deflating subspaces of the pencil $A - \lambda B$. Note that if B is invertible then we can choose $\hat{B} := B$ and $\hat{A} := A$.

8 Numerical experiments

We report on numerical experiments that illustrate the potential of the 2sGRQI method (Algorithm 3.1) as a numerical algorithm. The 2sGRQI method has been implemented in Matlab as described below.

Algorithm 8.1 (implementation of 2sGRQI) *Let C be an $n \times n$ matrix. Given two $n \times p$ matrices Y_{L_0} and Y_{R_0} satisfying $Y_{L_0}^H Y_{L_0} = I = Y_{R_0}^H Y_{R_0}$, the algorithm produces a sequence of matrices (Y_{L_k}, Y_{R_k}) as follows. For $k = 0, 1, 2, \dots$,*

1. *Compute the $p \times p$ block Rayleigh quotient $R_R := (Y_{L_k}^H Y_{R_k})^{-1} Y_{L_k}^H C Y_{R_k}$. Compute an*

eigendecomposition $R_R = W_R \text{diag}(\rho_1, \dots, \rho_p) W_R^{-1}$ using the Matlab `eig` function. Obtain the eigendecomposition $R_L = W_L^H \text{diag}(\rho_1, \dots, \rho_p) W_L^{-H}$ by computing $W_L := (Y_{L_k}^H Y_{R_k}) W_R$.

2. Solve the decoupled equations (3), that is, $(C - \rho_i I)(z_R)_i = Y_{R_k} W_{R_k} e_i$ and $(C^H - \rho_i I)(z_L)_i = Y_{L_k} W_L^{-H} e_i$, $i = 1, \dots, p$, using the Matlab “\” operator. If the solutions have any nonfinite element, then solve instead $(C - \rho_i I + \epsilon I)(z_R)_i = Y_{R_k} W_{R_k} e_i$ and $(C^H - \rho_i I + \epsilon I)(z_L)_i = Y_{L_k} W_L^{-H} e_i$, $i = 1, \dots, p$, with ϵ small (we took $\epsilon = 10^3 \mathbf{u} \|C\|_F$ where \mathbf{u} is the floating point relative accuracy and $\|C\|_F$ is the Frobenius norm of C).

3. Orthonormalize $Z_R := [(z_R)_1 \ \dots \ (z_R)_i]$ to obtain $Y_{R_{k+1}}$, and likewise for Z_L to obtain $Y_{L_{k+1}}$. In Matlab, orthonormalizations are performed using the “economy size” QR decomposition, `[YL,ignore] = qr(ZL,0)` and `[YR,ignore] = qr(ZR,0)`.

Note that if C , Y_{L_0} and Y_{R_0} are real, then the columns of Z_L and Z_R appear in complex conjugate pairs and unnecessary work can thus be avoided in the computation of Z_L and Z_R .

It is well known [BS89] that the basins of attraction of RQI (Algorithm 2.4) may collapse around attractors when the eigenvalues of A are not well separated. This property also holds for GRQI [ASVM04] and obviously extends to 2sGRQI (Algorithm 3.1). Moreover, in 2sGRQI the matrix C is not necessarily Hermitian; its eigenspaces can thus be arbitrarily close to each other. In a first set of experiments, in order to ensure a reasonably large basin of attraction around the left-right eigenspaces, we ruled out clustered eigenvalues and ill-separated eigenvectors by choosing C as follows: $C = SDS^{-1}$, where D is a diagonal matrix whose diagonal elements are random permutations of $1, \dots, n$ and $S = I + \frac{\alpha}{\|E\|_2} E$, where the elements of E are observations of independent random variables with standard normal distribution and α is chosen from the uniform distribution on the interval $(0, 0.1)$. The initial matrices Y_{L_0} and Y_{R_0} are randomly chosen such that $\text{dist}([Y_{R_0}], [S(:, 1:p)]) < 0.1$ and $\text{dist}([Y_{L_0}], [S^{-H}(:, 1:p)]) < 0.1$, where “dist” is the largest principal angle.

Algorithm 8.1 was run 10^6 times with $n = 20$, $p = 5$. The matrices C , Y_{L_0} , and Y_{R_0} were randomly chosen in each experiment as explained above. Experiments were run using Matlab 7.2 with floating point relative accuracy approximately equal to $2 \cdot 10^{-16}$. Results are summarized in Table 1, where the error e is defined as the largest principal angle between $[Y_R]$ and $[S(:, 1:p)]$ plus the largest principal angle between $[Y_L]$ and $[S^{-H}(:, 1:p)]$. These results show that convergence to the target eigenspace occurred in each of the 10^6 runs. The evolution of the error is compatible with cubic order of convergence.

Iterate number	mean(log10(e))	max(log10(e))
0	-1.4338	-1.0000
1	-4.6531	-2.6338
2	-13.9359	-8.3053
3	-16.5507	-15.1861
4	-16.5524	-15.1651
5	-16.5509	-15.1691

Table 1: Numerical experiments for Algorithm 8.1. See details in the text.

The behavior of the 2sGRQI algorithm in case of ill-separated eigenvectors/values would deserve investigation. The Hermitian case is studied in [ASVM04] where improvements of GRQI and the Riemannian Newton algorithm are proposed.

In another set of experiments, real Hamiltonian matrices C were selected randomly as

$$C = \begin{bmatrix} F & \tilde{G} + \tilde{G}^H \\ \tilde{H} + \tilde{H}^H & -F^H \end{bmatrix}$$

where F , \tilde{G} and \tilde{H} are matrices of dimension $\frac{n}{2} \times \frac{n}{2}$ whose elements are independent observations of the standard normally distributed random variable. A new matrix C was selected for each experiment. For testing purposes, an eigenvalue decomposition $C = SDS^{-1}$ was computed using the Matlab `eig` function, and the *full* left and right real eigenspaces corresponding to the eigenvalues with largest real part in magnitude were chosen as the target left and right eigenspaces. (The notion of *full* eigenspace is defined in Section 7.2. The *real* eigenspace associated to a pair $(\lambda, \bar{\lambda})$ of complex conjugate eigenvalues with eigenvectors $v_r + iv_i$ and $v_r - iv_i$ is the span of v_r and v_i .) The eigenvalue decomposition was ordered in such a way that $\lfloor S^{-H}(:, 1:p) \rfloor$ is the target left-eigspace and $\lfloor S(:, 1:p) \rfloor$ is the target right-eigspace. Note that we have $p = 2$ when the target eigenvalues are real (λ and $-\lambda$), or $p = 4$ when the target eigenvalues have a nonzero imaginary part (λ , $\bar{\lambda}$, $-\lambda$, and $-\bar{\lambda}$). The initial matrix Y_{R_0} was randomly chosen such that $\text{dist}(\lfloor Y_{R_0} \rfloor, \lfloor S(:, 1:p) \rfloor) < 0.1$, where “dist” is the largest principal angle, and Y_{L_0} was chosen as JY_{R_0} in accordance with the material of Section 7.2. Convergence to the target left and right eigenspaces was declared when the error e as defined above was smaller than 10^{-12} at the 10th iterate. Algorithm 8.1 was run 10^6 times with $n = 20$ and $p = 4$ with the matrices C , Y_{L_0} and Y_{R_0} randomly chosen in each experiment as described above. Note that, in accordance with the material in Section 7.2, only Z_R was computed at each iteration; Z_L was chosen as JZ_R . We observed that convergence to the target eigenspaces was declared for 99.95% of the 10^6 experiments. Next, the experiment was run 10^6 times with the distance bound on the initial condition set to 0.001 instead of 0.1. Convergence to the target eigenspaces was declared for all but seven of the 10^6 randomly generated experiments. This confirms the potential of Algorithm 8.1 for refining initial estimates of *full eigenspaces* of Hamiltonian matrices.

9 Conclusion

We have shown that Ostrowski’s two-sided iteration generalizes to an iteration on $\text{Grass}(p, n) \times \text{Grass}(p, n)$ that converges locally cubically to the pairs of spectral nondefective left-right eigenspaces of arbitrary square matrices. The iteration is competitive with Chatelin’s Newton method and it yields one-sided formulations adapted to some structured eigenproblems, including the Hamiltonian and generalized Hermitian eigenproblems.

Acknowledgements

This work was initiated when the first author was a guest in the Mathematics Department of the University of Würzburg under a grant from the European Nonlinear Control Network. The hospitality of the members of the Department is gratefully acknowledged. The first author would also like to thank especially Rodolphe Sepulchre for his careful guidance throughout his PhD research and the impact this had on the results of this paper.

References

- [AMS04] P.-A. Absil, R. Mahony, and R. Sepulchre, *Riemannian geometry of Grassmann manifolds with a view on algorithmic computation*, Acta Appl. Math. **80** (2004), no. 2, 199–220.
- [AMS07] P.-A. Absil, Robert Mahony, and Rodolphe Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, Princeton, NJ, 2007, in preparation.
- [AMSV02] P.-A. Absil, R. Mahony, R. Sepulchre, and P. Van Dooren, *A Grassmann-Rayleigh quotient iteration for computing invariant subspaces*, SIAM Rev. **44** (2002), no. 1, 57–73.
- [ASVM04] P.-A. Absil, R. Sepulchre, P. Van Dooren, and R. Mahony, *Cubically convergent iterations for invariant subspace computation*, SIAM J. Matrix Anal. Appl. **26** (2004), no. 1, 70–96.
- [BBMX02] Peter Benner, Ralph Byers, Volker Mehrmann, and Hongguo Xu, *Numerical computation of deflating subspaces of skew-Hamiltonian/Hamiltonian pencils*, SIAM J. Matrix Anal. Appl. **24** (2002), no. 1, 165–190.
- [BS89] S. Batterson and J. Smillie, *The dynamics of Rayleigh quotient iteration*, SIAM J. Numer. Anal. **26** (1989), no. 3, 624–636.
- [BS90] ———, *Rayleigh quotient iteration for nonsymmetric matrices*, Math. Comp. **55** (1990), no. 191, 169–178.
- [Cha84] F. Chatelin, *Simultaneous Newton’s iteration for the eigenproblem*, Computing, Suppl. **5** (1984), 67–74.
- [Dem87] J. W. Demmel, *Three methods for refining estimates of invariant subspaces*, Computing **38** (1987), 43–57.
- [EAS98] A. Edelman, T. A. Arias, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl. **20** (1998), no. 2, 303–353.
- [GH06] A. Galántai and Cs. J. Hegedüs, *Jordan’s principal angles in complex vector spaces*, Numer. Linear Algebra Appl. **13** (2006), no. 7, 589–598.
- [GLAM92] J. D. Gardiner, A. J. Laub, J. J. Amato, and C. B. Moler, *Solution of the Sylvester matrix equation $AXB^T + CXD^T = E$* , ACM Trans. Math. Software **18** (1992), no. 2, 223–231.
- [GLR86] I. Gohberg, P. Lancaster, and L. Rodman, *Invariant subspaces of matrices with applications*, Canadian Mathematical Society Series of Monographs and Advanced Texts, John Wiley & Sons, New York, 1986.
- [GV96] G. H. Golub and C. F. Van Loan, *Matrix computations, third edition*, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, 1996.
- [HS03] Michiel E. Hochstenbach and Gerard L. G. Sleijpen, *Two-sided and alternating Jacobi-Davidson*, Linear Algebra Appl. **358** (2003), 145–172, Special issue on accurate solution of eigenvalue problems (Hagen, 2000).
- [Ips97] I. C. F. Ipsen, *Computing an eigenvector with inverse iteration*, SIAM Rev. **39** (1997), no. 2, 254–291.
- [JS00] Zhongxiao Jia and G. W. Stewart, *An analysis of the Rayleigh-Ritz method for approximating eigenspaces*, Math. Comp. **70** (2000), no. 234, 637–647.
- [LE02] E. Lundström and L. Eldén, *Adaptive eigenvalue computations using Newton’s method on the Grassmann manifold*, SIAM J. Matrix Anal. Appl. **23** (2002), no. 3, 819–839.
- [Ost59a] A. M. Ostrowski, *On the convergence of the Rayleigh quotient iteration for the computation of the characteristic roots and vectors. I-VI*, Arch. Rational Mech. Anal. (1959), 1, 233–241, 2, 423–428, 3, 325–340, 3, 341–347, 3, 472–481, and 4, 153–165.
- [Ost59b] ———, *On the convergence of the Rayleigh quotient iteration for the computation of the characteristic roots and vectors. III*, Arch. Rational Mech. Anal. **3** (1959), 325–340.
- [Par74] B. N. Parlett, *The Rayleigh quotient iteration and some generalizations for nonnormal matrices*, Mathematics of Computation **28** (1974), no. 127, 679–693.

- [Par98] ———, *The symmetric eigenvalue problem*, Classics in applied mathematics, vol. 20, SIAM, Philadelphia, 1998.
- [PW79] G. Peters and J. H. Wilkinson, *Inverse iteration, ill-conditioned equations and Newton's method*, SIAM Review **21** (1979), no. 3, 339–360.
- [Sch01] K. Scharnhorst, *Angles in complex vector spaces*, Acta Appl. Math. **69** (2001), no. 1, 95–103.
- [Smi97] P. Smit, *Numerical analysis of eigenvalue algorithms based on subspace iterations*, Ph.D. thesis, CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands, 1997.
- [Ste73] G. W. Stewart, *Error and perturbation bounds for subspaces associated with certain eigenvalue problems*, SIAM Review **15** (1973), no. 4, 727–764.
- [Ste01] G. W. Stewart, *A generalization of Saad's theorem on Rayleigh-Ritz approximations*, Linear Algebra Appl. **327** (2001), no. 1-3, 115–119.
- [Wil65] J. H. Wilkinson, *The algebraic eigenvalue problem*, Clarendon Press, Oxford, 1965.